UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460

October 7, 2003

OFFICE OF
THE ADMINISTRATOR
EPA SCIENCE ADVISORY BOARD

Note to the Reader:

The attached report is the first draft of the EPA Science Advisory Board (SAB) 3MRA Panel. This draft responds to charge questions 1, 3 and 4; the response to question 2 will be provided separately when available. The draft will be discussed at the October 9 public conference call meeting of the Panel. This draft has not been reviewed by the Panel and cannot be considered as the consensus position of the panel involved in the review. Additional drafts will be developed and considered. Once a draft is approved by the Panel, it will be considered bu the Executive Committee. Only after been approved by the EC, will the report be transmitted to the EPA Administrator and become available to the interested public as a final report.

This draft has been released for general information to members of the interested public and to EPA staff. The reader should remember that this is an unapproved working draft and that the document should not be used to represent official EPA or SAB views or advice. Draft documents at this stage of the process often undergo significant revisions before the final version is approved and published.

The SAB is not soliciting comments on the advice contained herein. However, the EPA Program Offices requesting the review may respond to the issues listed below at this time or in response to a later draft.

1. Has the Committee adequately responded to the questions posed in the Charge?
2. Are any statements or responses made in the draft unclear?
3. Are there any technical errors?

For further information or to respond to the questions above, please contact:

Kathleen White, Designated Federal Officer
EPA Science Advisory Board (1400A)
US Environmental Protection Agency
1200 Pennsylvania Avenue, NW
Washington, DC 20460-0001
(202) 564_4559 Fax: (202) 501-0582
E-Mail: white.kathleen@epa.gov

Charge Question 1. While the EPA had the assessment methodology peer reviewed prior to the development of the 3MRA modeling system, does the SAB have any additional comments about the methodology as implemented?

The 3MRA assessment methodology contains several elements:

- Statistical sample of industrial sites
- Site-based human and ecological exposure/risk assessment
- Multi-contaminant, -media, -pathway, -receptor
- Tiered Data (site-specific, regional, national)
- Population-based site level risk estimates
- National roll-up of risks
- Alternative measures of protection
- Two-stage Monte Carlo
- Probability-based design to facilitate uncertainty analysis and sensitivity analysis
- Externally peer reviewed and independently tested

The panel concurs that the 3MRA modeling system is a major step forward in providing a screening level, computer-based tool for estimating the distributions of risk and hazard that result from various choices of exit threshold, and provides a scientifically defensible basis for determining exit levels for RCRA-listed hazardous wastes. Overall, it is clear that every attempt was made to choose the optimal design for each module mindful of the balance between the computing feasibility and the physical realism achievable by the simulation. The documentation presents evidence that the designers tried to follow the main features of the assessment methodology, and were mindful of the criticisms of the HWIR 95 version, especially the requirement to conserve mass, the need for a true multi-media, multi-pathway, multi-receptor approach, the necessity of devising a meaningful validation approach, and the value of transparent documentation.

The panel also commends the manner in which 3MRA was developed, i.e. as a genuine cross-Agency effort that to a significant degree worked through the insular nature of individual units in a large organization, forming a formal partnership between the Office of Solid Waste and the Office of Research and Development. This trend shows evidence of continuing as work on comparisons of model output between 3MRA and TRIM-fate, developed independently by the Office of Air Quality, proceeds. In addition, adherence to the principles articulated by the Committee on Regulatory Environmental Modeling (CREM) by the 3MRA team is clear. These approaches serve as excellent examples of how Agency leadership and inter-Office cooperation can effectively serve the greater needs of the regulated community and the public.

The panel recognizes the rationale of a tiered set of data for conducting screening level assessments, and the use of statistical sampling and analysis that together define the approach for developing a national assessment methodology. However some panel members express concerned about the adequacy of the site-specific database for conducting national assessments by waste management type, about the use and interpretation of the pseudo two-dimensional Monte Carlo analysis for assessing variability and uncertainty,

and about the EPA's response, or lack thereof, to independent reviews of individual modules of 3MRA.

The 3MRA system is intended to rest on sound scientific principles, among them the conservation of mass. The model is designed to conserve mass within each waste management unit, and within each legacy model, allowing for partitioning among gaseous, liquid, and solid phases. However, it is unclear that mass is conserved at the point of "handoff" between modules. Probably the only way to confirm mass conservation is to assume that the legacy fate and transport models internally conserve mass and to check the flows from one another. For example, mass conservation in the food chain portion of the exposure assessment can be assured only if there are feedback loops that subtract partitioned COPC (what is this?) quantities from the contact medium. COPC taken up by aquatic biota should be subtracted from that in the water; mass removed from the air by air-to-plant transfer should be subtracted from that in the air. These are cul de sacs for which mass accumulations can become significant for high values of the partition coefficients. Each of these corrections will depend upon a running tally of the standing crop (of fish, plants or whatever is inducing intermedia transfers). Fugacity models automatically maintain such mass balances, but assemblages of dynamic fate and transport models such as that in 3MRA do not. Since mass conservation is such a fundamental aspect of model development, the panel is concerned that a tangible demonstration be carried out under conditions that would provide a stringent test, i.e. using different classes of contaminants that partition strongly from one phase to another.

The panel is concerned about the lack of sophistication, in comparison with transport and fate, of the treatment of risk in 3MRA. Probably the largest degree of uncertainty in both the human health and ecological risk assessment protocols is associated with dose-response relationships. Notwithstanding Agency policy to the contrary, the panel feels that the Monte Carlo analysis should recognize these uncertainties as well as species response variabilities. Probability density functions characterizing cancer slope factors, reference concentrations and reference doses should be derived from the original toxicological and/or epidemiological databases. The panel strongly endorses the inclusion of such an approach into 3MRA as future versions are developed. In the meantime, the 3MRA documentation should be very clear on the meaning of the risk and hazard estimates corresponding to the exit levels. Statements about the model output should be carefully crafted to avoid the impression that the selected percentiles of protection are accurate representations of actual population protection. A suggested interpretation of the output protection levels: "For (a given %) of the WMUs in the U.S. there is no more than a (n%) chance that an individual - selected at random from the population within (0.5, 1.0, or 2.0) km of the WMU – would be exposed to a long-term average concentration of chemicals originating from the waste that is associated with an upper-bound estimate of lifetime extra cancer risk that equals or exceeds one chance in (one million, one hundred thousand, etc.). Actual risk is probably less than the upper-bound estimate and may be zero."

Exposure duration is also input as a constant value in 3MRA, neither variable nor uncertain. This fixed value fails to capture the upper end of the distribution. Farm families in particular often spend their whole lives at one residence. Again, this lack of variation and possible underestimate for

farmers distorts the output distribution. : Furthermore, several exposure parameters that co-vary with body weight are treated as independent. This makes the exposure appear more variable than it really is. For example, respiration rates are not normalized to body weight, nor are they apparently correlated to body weight. This means that body weights and respiration rates are selected randomly and independently from their distributions for each model realization. The result of this approach is that the largest adult (660 lbs) is just as likely to be paired with the minimum breathing rate (1 m^3/day) as with the maximum breathing rate (50 m^3/day). This is also true for the smallest adult (33 lbs). This means that the breathing rate can cover the implausible range from 3.3 L/kg/day to 3,300 L/kg/day. Were individual iterations of the model stored and the tails of the distribution studied, such pairings will probably be found. Ideally, respiration rates would be expressed as a function of body weight to the 0.7 power. Other examples of variables incorrectly treated as independent of body weight include fish consumption and drinking water consumption. Again, the problem is similar: when the exposure to contaminants in the fish or in drinking water is expressed in mg/kg/day, the range of exposure rates is exaggerated because the model allows, say, a 10 kg child to eat 1500 g fish or drink 2100 ml of water per day. At a minimum, respiration and drinking water and fish consumption should be normalized body weight or at least correlated with body weight.

3MRA and the framework in which it is employed make no distinction between a waste that contains a single regulated chemical at a concentration just below the exit level and a waste that contains several regulated chemicals each at concentrations just below the exit level. A simplistic but effective solution to the problem of possible additive effects among multiple chemicals in a waste stream is to require that the concentration of each regulated chemical in the waste stream be expressed as a fraction of its exit concentration and that the sum of these fractions be no greater than one (1).

Although the risk characterization within three MRA is said to offer individual risk distributions, in the general case these consist of three end points: two for humans (cancer, non-cancer over a lifetime exposure), and one ecorisk based on the calculation of the hazard index for sensitive organisms (which is interpreted as the reproductive impairment of a mating species). There is no consideration of "non-lethal" endpoints for humans (e.g. formal disability), or overall ecosystem structure and functioning that might induce human risks (e.g. the collapse of an ecosystem "service" upon which humans depend). It is argued that limited availability of data, lack of general models, and sensitivity to site-specific aspects render system level ecological treatment difficult if not impossible to implement at this time. This suggests that ecosystem dynamics modeling be made part of the supplemental science-based

initiatives. Part of this research should focus on the treatment of non-linear terms in aquatic ecosytem dynamics, such as competitive mortality (like the Lotka-Volterra formalism), and the feeding terms that are the products of biomasses of feeders and foods. These terms are further influenced by chemical contaminants that inhibit energy flows and interfere with the population dynamics at various specific trophic levels. With data generated from past and future field research, the developers should fashion generic biome-related ecosystem models to avoid having to generate separate ones for each site as is necessary with the met data. A deliberate effort in this direction will immediately indicate the gaps in our knowledge thereby permitting useful feedback to the field research.

Perhaps the most complex issue that the panel has faced in evaluating the 3MRA modeling system has been that of validation. 3MRA is a complex higher order model that does not lend itself to traditional methods of validation, i.e. in the sense of data matching. While such an approach can be achieved for some of the model components, such as waste management unit and fate and transport models, it is not possible to perform such a validation on the model as a whole for two reasons: because a complete dataset that stresses all seventeen of the sub-models simultaneously does not exist and is unlikely to become available soon, and because, ultimately, the purpose of 3MRA is to perform a national risk assessment. The Agency's approach to this has been to develop a tiered validation protocol, based heavily on the work of Beck et al. (1997). In this scheme, validation is seen as a design problem with several elements:

- Quality of input data (volume 2 of the 3MRA material)

- Quality of model components (volumes 1 and 3)

- Quality of the modeling system (also in volume 3)

- Performance of the model as a reliable instrument for its assigned task (performance validity). Uncertainty and sensitivity analysis are central to the concept of performance validity, as is comparison with other models

(e.g. TRIM fate), and matching against available but limited datasets (a chlor-alkali site). These are the subjects of volume 4.

Validation is achieved through completion of a series of well-defined tasks that must meet rigorous quality assurance evaluations of their outcomes. This approach represents a shift away from equating model validity with its ability to correctly predict the future, a future that in a policy context is fundamentally unknowable, to a focus on the quality and reliability of model forecasts (minimum risk of an undesirable outcome).

In this context, the Agency has described in detail the problem that needs to be solved (national risk assessment), has designed a method for obtaining a solution (the 3MRA risk assessment methodology), and has generated the "solution" (the 3MRA model system). At present they are in the early stages of evaluating the performance validity of the modeling system for generating reliable forecasts. Thus in terms of the steps above, they have accomplished the first three and are engaged in the fourth.

It is clear to the panel that in each of the stages of model validation the Agency has set forth extensive quality assurance procedures that include consensus on the model's intended use and performance criteria; incorporation, whenever possible, of legacy models with which the scientific community has considerable experience; independent peer reviews of model architecture and components; and verification of computer code and inter-model communication. Thus in evaluating 3MRA, the panel has had to first grasp the basis of the validation protocol, and then assess the degree to which the Agency has achieved what it set out to do. The issues raised above are addressed in detail in charge questions 2, 3, and 4 below.

In addition to implementation of the assessment methodology, charge question 1 also asks if the panel has any additional comments about the 3MRA assessment methodology. As it presently stands the scope of 3MRA is limited to developing a tool to help make exit decisions, in terms of chemical concentrations, on the migration of waste streams from RCRA subtitle C to "ground based" subtitle D facilities, referred to as waste management units (WMUs). As such, it excludes obvious allocation alternatives by adopting a conservative, and distinctly 20th century view, that encompasses a limited range of final disposal

options. This grows from the dependence on the inventory of candidate facilities dating back in the decade of the mid 1980s, and reflects the underlying influence of the Resource Conservation and Recovery Act on the motivation for developing 3MRA. Indeed, the Agency has embarked upon a thought process to reconsider the basis and procedures of RCRA to make it more congruent with its original goal to encourage recycling and reuse of materials ("Rethinking RCRA" white paper reference).

The choice of focusing only on land-based facilities has some intrinsic limitations; for example, the landfill prototype and the surface impoundment WMU are put down on native material without the benefit of liners, gas collection systems or leachate treatment systems. It is unclear to the panel how many facilities fitting these descriptions are still allowed to operate even considering the range of regulatory oversight under state jurisdictions, but in any case the panel feels that some representative range of modern technology should become available to the 3MRA user. This becomes especially important as the uses of 3MRA are extended to include evaluation of individual facilities (for example for delisting purposes), or designs for proposed facilities. In addition, the consequence of allowing relatively unsophisticated protective designs is the influence on exit levels, which may turn out to be unrealistically restrictive thereby defeating one of the major purposes for developing 3MRA, that of overly stringent regulation. In order to treat the case of a modern Subtitle D landfill, there will need to be sub-models to determine gas generation, collection and utilization, contaminant levels in fugitive emissions of uncollected gas, pollutant partitioning in cover soil, and failure mode analysis of these processes.

The panel's view is that the present assessment methodology misses out on at least four strategies for releasing a waste stream from the rigors of Subtitle C; they are: support for delisting of hazardous wastes, municipal waste combustors, and pollution prevention and industrial ecology alternatives. By omitting such options, the 3MRA assessment methodology needlessly restricts the decision-maker's thinking by offering only the five classes of WMUs included in the simulation, when in reality, the missing alternatives are readily implemented and officially encouraged under available contemporary practices. These alternatives are amplified below.

An immediate application of 3MRA would be to support de-listing petitions. For this use it needs to be set up in such a way that site-specific data can be readily entered to supplement the existing databases, and enough iterations run for a single site

to give reproducible results. Beyond this, there is potential for assessment of risk at contaminated sites, and risk-based support of permitting decisions. Because the FRAMES architecture allows for plug-in applications to suit specific needs, it can have many other uses. The panel supports the use of this model with some enhancements for the intended purpose and its continued development for other applications.

Considering that, after materials recovery, 21% of US municipal waste is handled by municipal waste combustors (MWCs), it is surprising that this diversion alternative is not included in 3MRA. Preliminary studies have suggested the favorable feasibility of destroying household hazardous waste in MWCs considering the temperature-time characteristics of MWC furnaces Very effective destruction and removal efficiencies are available in the MWC even for such refractory compounds as CFCs. In its present form, the 3MRA modeling system has all of the modules needed to assess risks of air emissions and ash disposal from MWCs, and sufficient data exist to support a combustion alternative that can be evaluated on a national scale. Emission rates and configuration parameters (e.g., stack characterization: height, diameter, temperature, velocity and base elevation) are available for the US population of MWCs, although the range of receptor domains needs to be enlarged for each source because of elevated (meaning higher altitude?) emissions. All of the algorithmic mechanisms for handling deposition and indirect pathways are already embodied the present version of the 3MRA so that only the source and receptor files need to be modified.

A second type of WMU alternative that 3MRA might address are pollution prevention schemes involving stabilization of a hazardous waste in a product stream. An example of this is the exemption of petroleum coke quenched with oily refinery sludges from the standards, record keeping and labeling requirements of RCRA. Since the early 1970s some refineries have blended API separator sludges, tank bottoms and biological solids in the water stream used to cool petroleum coke at the end of a delayed coker cycle. Presumably contaminants such as metals or polycyclic aromatic hydrocarbons bind to the carbonaceous substrate of coke particles. This technique has been embraced by the European Commission in its catalog of Best Available Technologies, but it is doubtful that any occupational or community health risk assessment was ever performed (reference?). Evaluation of exemptions such as this should have a clear place in the 3MRA assessment methodology. A generic module with adjustable input/output structure might be contemplated in anticipation of problems like this. As further encouragement to the user, some synthetic case studies might be packaged in with the software as a means of demonstrating the flexibility of 3MRA.

The third example is taken from the industrial ecology field: the use of a waste stream from one process as a raw material for another. This example grows out of a toxic tort complaint under California's Proposition 65. A tannery cleaned sheepskins with perchloroethylene prior to using the skins to make high fashion shearling coats. The operator filed a Form R for the Toxic Release Inventory listing all of the makeup perc (thousands of pounds per year) under air releases. An advocacy group immediately brought suit under Proposition 65 that there existed significant risk and that the operator should have posted a notice. After a careful mass balance on the plan, it was discovered that nearly all of the makeup perc was shipped out the front door in lanolin, a by-product of the process. Had this perc been separated, it would have been classified a hazardous sludgy waste. The Proposition 65 action was successfully defended in a court trial, but no risk assessment was done on the subsequent uses of the lanolin because there was no regulatory basis for it. One might argue that the recipient of the perc-contaminated lanolin had a cause of action, but that never came to pass and the tannery moved to India. This is a case where a flexible access to the 3MRA modeling system could be ported to accept generic inputs for a myriad of alternative waste diversion strategies. As in the case of pollution prevention, a synthetic demonstration study will encourage users to consider other possible waste management scenarios.

[NOTE: The integrated response to Charge Question 2 is not available at this time. It will be transmitted separately as soon as it is available.]

Charge Question 3a: Is the software development and verification testing approach implemented for the 3MRA modeling system sufficient to ensure confidence that the modeling results reflect the modeling system design?

Interpretation

This question asks whether the 3MRA modeling system code implements a quantitative calculation that is consistent with the model conceptual design and whether EPA has "verified" that the code computes what it is intended to compute. This question is interpreted as dealing with the "correctness" of the 3MRA model code. Ideally, one would like to test the code against an analytical solution of the same algorithms. For a modeling system this complicated however, this test is not possible for the entire model; although, the approach of comparison with analytical solution might be valuable for individual modules or processes within modules. Also, the question of verification may have some overlap with Question 2 in that the code may be "verified" to be accurately reproducing the computations inherent in the model conceptual design but that conceptual design may have flaws in basic assumptions or process formulations. Verification also refers to proper QA/QC in code and database development including proper maintenance and documentation of modifications to correct errors, and full testing across all aspects of the modeling system's functionality.

Response to Charge Question 3a

General Comments

The EPA has made a reasonably good effort, especially in implementing the process described in Volume 3-Section 3.1, of verifying that the 3MRA functioning according to its design. The detailing of the depth and breadth of the carefully designed functional testing routines and the multitude of test runs performed lead the reviewer to a sense of confidence that the only random events occurring in model runs are those programmed into the Monte Carlo analyses. The expanded documentation on the structure of the entire 3MRA Modeling System superstructure and content should help all reviewers gain confidence in the soundness of the overall approach to computer model development. The special attention given to the development team communication and "top-down code design", as well as conduct of QA/QC testing according to a pre-planned testing strategy, are particularly notable. Also, individual modules have been verified; e.g., EXAMS I/O has been tested to assure that it is

working properly in transferring data with other modules. The use of the FRAMES foundation also appears to be a plus for making the model more adaptable to future modifications, with less repetition of structural testing.

Specific Comments

One reviewer (Boissevain) expressed a concern that the data consolidation in the presentation of risk/HQ values may be underestimating the risks associated with an exit level. This issue of data consolidation for expediency should be verified with respect to its impact on risk.

Another verification issue (Brown) is related to the quantification of biases in model results based on the propagation of module assumptions/limitations (i.e., process/loading assumptions, module structure, etc.) through the system. Limitations and potential biases of individual modules have been qualitatively described for each module and in some cases the direction and rough magnitude is presented. It would be desirable to attempt to make estimates of biases for all modules more quantitative. Also, it is important for model developers to estimate how the module biases are propagated through the integrated system. In other words, does the bias inherent in the known module limitations tend to accumulate (positive or negative direction) or do they tend to offset one another? Can the extent of this accumulation/offset be estimated? If so, what would it be based on?

Summary

Historically, many environmental modeling codes have not had this level of attention to the fundamental quality of the code structure and performance testing. It would appear that this phase of the model testing is reasonably complete and that most significant need for further attention to model quality will fall in the area of "field validation" and "sensitivity and uncertainty analysis" testing addressed in the second part of this charge, as discussed below.

<u>Charge Question 3b</u>: EPA has implemented thorough evaluations using the available data resources and technologies, while also recognizing the real world limitations that apply to validating the 3MRA modeling system. Have we reasonably demonstrated through methodology design, peer review, quality control, sensitivity analyses, and model comparison, that the 3MRA modeling system will produce scientifically sound results of high utility for use in multi-media regulatory applications?

Interpretation

Assuming the reply to question 3a is affirmative, this question deals with the ability of the 3MRA modeling system to reproduce actual conditions relative to risks to human and ecological receptors caused by chemical releases from WMUs. In other words, can decision-makers confidently use this model to inform decisions regarding contaminant-specific exit levels for low-risk wastes. This question overlaps somewhat with question 2 of the charge, because its answer depends on whether the modeling system employs all the necessary processes for computing the quantitative relationship between the operation of a WMU and the local risks that its operation poses; and it depends on whether those processes have been represented mathematically in a scientifically sound and up-to-date manner. But the question also deals with the integration of those processes, the temporal and spatial scales at which they operate, and the parameterization of those processes relative to the time and space scales at which they are operating. In short, the question asks whether the activities to date have produced a modeling system that is sufficiently validated for use in regulatory applications for which it was intended.

In addressing question 3b, we must recognize the reality that fully "validating" an extremely complex model representation of a natural system is virtually impossible. The best that one can hope for is to conduct enough confirmation of the model to build sufficient confidence that it computes the quantitative response of the system to external forcing functions (including chemical loads and other important stressors) within a stated level of tolerance (acceptable uncertainty)? In other words, do we have confidence that the model can inform the defined management goal – determining chemical-specific exit levels? If the model passes this test, then it is what I call "confirmed" as opposed to "validated". Because of the precedent set in this document, we will continue to use the term "validate"; but we will recognize that complete validation in the context of comparing the model to one or more coherent and independent data sets relative to its stated purpose has not been accomplished and probably will not be accomplished in the near future.

Given the above philosophy, it will be important to try to state a priori what model Performance Indicators (measures/metrics of how well the model simulates the cause-effect relationships of management concern) can be used to assess the validation of the

3MRA modeling system. The following response to this charge question will therefore be categorized by the following types of model Performance Indicators:

1. Do the conceptual model and module/process formulations in the 3MRA modeling system provide a scientifically valid representation of the system and problem of concern?

2. Are the spatial and temporal scales in the integrated model (or individual modules) appropriate for the problem context and data availability?

3. Does the model compare well with field observations of the cause-effect relationships of management concern? (The desired level of comparison (error tolerance) needs to be specified to apply this metric.)

4. Do the parameterization and computed rate of processes compare well with laboratory or field process experimentation?

5. Do the 3MRA model simulations compare well with theoretical expectations (e.g., mass balance, inter-media chemical distribution/partitioning, relative risks for a given chemical for various receptors and pathways, relative risks for a given receptor/pathway for various chemicals)?

6. Does the 3MRA model compare as expected with other models for a similar problem domain (e.g., multi-media or individual module (i.e., sub-model)) based on the respective spatial, temporal and process formulations of the two models?

7. Has the 3MRA modeling system or individual modules undergone previous peer review and been revised accordingly to address identified problems?

8. Does the 3MRA modeling system produce forecast simulations with sufficient accuracy and precision to

support exit level decisions? This analysis will rely on the sensitivity and uncertainty analyses that have been conducted with the model.

Response to Charge Question 3b

General Comments

Overall, the incorporation of a large number of "legacy" models into the structure of the 3MRA modeling system helps to "guarantee" that the methodology should be considered reliable and, on the whole, acceptable to the regulatory community and its various stakeholders. Most of these legacy modules have been independently developed, reviewed, tested, and constructively criticized by many researchers. They have also been frequently applied to practical regulatory permitting decisions in the current regulatory environment. Therefore, they are "familiar" to the risk assessment and environmental exposure modeling community. This familiarity brings with it an increased acceptance of their utility, but that same experience has produced lists of technical concerns about both inherent limitations in some of these modules, and about the uncertainties created by their application to situations in which their underlying assumptions may be violated or severely challenged. This experience with the individual models, and with their use in modeling systems different from 3MRA, sponsors the guarded comments raised by many in the public sector. However, individual module testing and confirmation does not guarantee that the integrated system of modules generates a computation that can support management decisions within the desired level of accuracy. In addition to the known assumptions and site-specific uncertainties associated with individual modules, I/O functions between modules that operate at different time/space scales may produce incompatibilities that lead to errors in overall cause-effect relationships and risk estimation.

As the model documentation in volume 3 recognizes, data-based confirmation of 3MRA is very difficult because of a scarcity of observations of risk against which to compare model predictions. The large scale of regional and national risk computations and the very long integration times inherent in risk calculations exacerbate this effort. For this reason, the model "validation" process should include all eight of the model performance indicators listed above. With respect to the conventional model validation metric (no. 3 above – comparison with field data),

the best solution is to confirm the model on as many individual areas of concern as possible covering a wide range of conditions relative to chemical types, WMU types, receptor conditions, and pathways of exposure. This should be an ongoing process as data sets of opportunity or resources for implementation of model validation field programs become available. Also, the field testing process should use a MC approach for model testing against the observed distribution of site data. In this way, an estimate of confidence level associated with a model result can be incorporated into the process.

Another general comment that relates to all categories of model performance indicators is the lack of consistency as to the meaning of qualitative descriptors of a given validation exercise. For example, Volume 3, Section 4.5 summarizes the verification /validation for EXAMS. It provides both quantitative and qualitative descriptors of the quality of the validation. Page 4-27, Cousin, et al. (1995) "Fairly good correlations were achieved between measured and predicted dissolved water concentrations, with predictions falling within a factor of 2 of station means". And, Armburst et al. (1999) " Predicted water column concentration responses generally matched the observed data. EXAMS overestimated soil concentrations by factors of 2 to 4." Here "fairly good" seems to mean predictions within a factor of 2 3,or 4, of observations.

On the other hand, the EPACMTP module validation section (Vol. 3, page 4-33) uses phrases such as "Demonstrated reasonable agreement.", "Accurately predicted…..", and "Demonstrated qualitatively similar results." There is no quantitative datum by which to compare these phrases to say the ones used for EXAMS. So how do we judge whether EXAMS is better or more poorly validated than EPACMTP, or any of the other modules? We need this type of module comparison so that we have some idea of where the "weak links" in the overall modeling system lie.

We suggest that there be a consistency in the use of qualitative descriptors for module validity. The phrase "reasonably good agreement" should mean the same thing (either in an absolute sense, say % difference, or a relative one, say best in class) when applied to the quality of the validation of different modules. It would also be useful if a quantitative mapping of what "reasonably good agreement" means were suggested, e.g. "reasonably good" means agreement within a factor of 2-4, as in the EXAMS reference cited above.

1. Scientific Validity of Conceptual Model and its Formulation

Issue of WMU source computation

Several of the SAB (Merrill, DePinto, others) have raised a question regarding how the WMU "source" term is used to model WMU chemical releases. From the panel discussions with EPA on 9/16/03, there remained some open questions regarding the manner in which the initial source concentration term is calculated for the WMU. Using the example of a Land-Based Source term, it would appear straightforward that the mass entering the WMU (for any particular waste application) is given by:

where:

C_w = concentration of chemical in waste (e.g,., mg/kg for solids)

f_{WMU} = fraction of waste stream in WMU (unitless 0.001 – 1)

ρ_b = bulk density of waste solids (kg/m^3)

V_w = total waste solids volume applied (m^3)

Yet, the source and transport modules are based on the chemical concentration (not mass) in the WMU to partition the chemical into its various media phases (see for example Equation 5-1 of Vol 1 for the Land-based source modules). The question for EPA is whether the concentration term applied to the source modules (sticking with the Land-Based example here) is the concentration in the original waste stream (C_w), or whether the source modules compute an "effective" initial concentration in the WMU that would be a function of the f_{WMU} term:

In this formulation the "effective" chemical concentration in the WMU is based on the chemical mass applied as a mass fraction

of the total WMU solids mass applied, giving an effective average chemical concentration for the WMU (recognizing that in a more rigorous statement, the total WMU solids mass can include an expression also expressed as a function of f_{WMU}). Note that an effective initial concentration could be calculated in this manner for each waste application event for the WMU.

While this may be an oversimplified depiction of the details of the source term, the basic question is whether the source concentration term is C_w or instead an "effective C_o" that takes f_{WMU} into account.

GSCM development

Dr. Thibodeaux has raised questions regarding the development and use (and associated uncertainties and limitations) of the GSCM (Generic Soil Column Model) and whether it represents a consistent use of science.

The GSCM was commissioned by the EPA because no such tool existed for the land-based waste source modules. There was a need for a generic, soil column, chemical fate and transport (CfaT) model that could serve multiple purposes in the 3MRA system. Such a general model could also be used in the soil of the watershed in addition to the waste in the WAUs. That was a very good concept; the 3MRA model developers were wise in initiating such and effort. Generally what was desirable from such a model was a science based algorithm that would mimic the key process features occurring in the soil column and then predict chemical volatilization to air, surface soil concentrations, and leachate quantity and quantity in the vadose among other things. The developers began with a pesticide soil fate model referred to as the Jury model. The GSCM description is contained in one section fo the final report (USEPA 1999. Source Modules for Non-wastewater WMUs (LAUs, WP, and Landfills) Background and Implementation for 3MRA for HWIR 99. OSW, Wash. DC October). The other sections were the application and implementation of th GSCM to all three of the WMUs and the watershed.

A review of the entire document was performed in Dec.1999 (ERG 1999). Six reviewers were used; they are listed on page 4-8. I read the commissioned model final report and the reviewers comments. Concerning the GSCM and the key role it plays in the entire 3MRA system I am uncomfortable with the level of external review it has received. The GSCM is the key piece in three of

the five source WAU modules in addition to being the basis for the watershed modules down gradient of the WP and the LAU. It therefore is a very important piece in the system and the reason it was given the named generic. The original reviewers were well qualified and did a good job, generally. The GSCM deserved its individual review in my opinion and still does. The reviewers raised numerous concerns about the mathematics/solution/status of implementation of the CfaT equation. For example, at the time of the review there were no numerical results available to the reviewers and the developers were in the midst of implementing the quasi-analytical solution approach. In reviewing the GSCM in much detail I have several concerns about its appropriateness for the HWIR screening level objective. In short I do not believe it is a good chemical process release model as implemented in the mathematically algorithm and it is not conservative. There are some very creative and innovative features in GSCM. Although it contains the right elements it is implemented so as to maintain a mathematical solution rigor at the expense of flexibility that would allow more process realism. In my opinion the mathematical solution is "over the top" and places restrictions on the model that are not necessary. A model consisting of fixed number of completely mixed soil layers is an alternative that deserves consideration in my opinion. I have a list of comments, concerns and suggestions that is more detailed and not appropriate at this particular time.

As a stand-alone process algorithm the GSCM should have been validated as a single entity. The basic process models used in the Surface Impoundment Module and the Aerated Tanks Module have had a long history of use and verification studies. The GSCM validation was performed within the context of validating the Land-based Source Module and Watershed Module (Section 4.2, Volume 3.). It's validation and that of the modules is the subject of the next section.

In summary, I believe that the GSCM is not ready for the key role it plays in the HWIR screening level task. It is a very new and innovative model; it was not finished at the time it was reviewed; it was one part of a set of other very important modules in the review package; many concerns were expressed by the reviewers about the model and some of those issues remain in the present version; the GSCM has not under gone an independent validation.

Other science issues

One basic question raised by a number of reviewers is whether the underlying biology and toxicology is appropriate to the objectives of this system. This question deals not only with the scientific validity of process formulations but with the question of whether they are appropriately applied and properly integrated within the whole system. Some concerns regarding the underlying science (these may be dealt with in more detail in other charge question responses) that have been raised are:

- Section 4.6.3 (Protection Criteria) appears to suggest that protection criteria can or may be selected for individuals, populations, and sites. The details of this approach are not clear, but protection may be inadequate where it is based on the product of multiple layers of protection criteria. For example, development of an HQ for aquatic life based on ambient Water Quality Criteria incorporates one level of protection criteria (WQC are designed to protect less than 100% of the species in the database less than 100% of the time), which then appears to be layered upon (or within) another level of protection criteria at the site level (e.g., 90 or 95% levels of protection). As a result, application of exit levels at the 90% level of protection for sites may result in species-level protection well below 90% (and only for those species in the WQC database).

- Use of the MATC in development of ambient Water Quality Criteria for aquatic life incorporates additional reductions in levels of protection. For example, a comparison of the MATC with the EC25 (effective concentration at which 25% of test organisms exhibited an adverse response) showed that the average MATC/EC25 ratio was greater than 2.0; thus, use of MATCs to derive HQs for aquatic life incorporate additional, hidden reductions (or at least uncertainty) in protection at the species level. Detailed analysis should be conducted to determine the risk posed to populations of aquatic (and likely other) organisms as a result of these multiple layers of protection criteria.

- Ecological risks are also calculated without assessment of the risks and impacts associated with concurrent exposure to multiple contaminants and multiple non-chemical stressors. Both of these limitations are acknowledged in the module documentation, but both pose significant limitations on the validity of the model. Further analysis of these limitations is necessary.

- Dermal exposure is not considered in 3MRA, yet significant efforts at EPA and elsewhere have been conducted since 1995 to assess and predict dermal exposure and its effects (its contribution to aggregate or cumulative exposure and risk).

- It is unclear how exposure aggregation occurs over different time periods (e.g., ages). Is risk calculated for each exposure (age) period independently, or is risk cumulative over several exposure periods. Clarification is necessary here.

- The documentation appears to discount cancer risk where exposure occurs only for a portion of the lifetime. However, some chemicals such as vinyl chloride pose a lifetime risk even if exposure is for less than a full lifetime. Clarification is required to determine how less-than-lifetime exposures to compounds such as VC are incorporated in the model.

- Exposure and risk is dependent on the time series of "available" chemical concentrations in media to which a given receptor is exposed. For example, does the use of annual average conditions for flow and other forcing functions like temperature, lead to erroneous exposure concentrations?

- In a similar vein, the assumption that TSS is modeled as a conservative substance in EXAMS – solids deposition and burial are not modeled – leads to a question of how a hydrophobic substance exchanges accurately with sediments and is lost to the exposure pathway by deep burial.

2. Model Spatial and Temporal Scales

Site-specific parameterization will depend on matching time and space scale of site-specific data with time and space scale of model process being parameterized. Model parameters are often very time/space scale dependent. This raises an important general question about the modeling system. Have some of the choices of regarding model complexity (as determined by spatial, temporal, and kinetic resolution) led to unnecessary and perhaps inappropriate "simplification" in order to accommodate two other 3MRA modeling features: (1) length of time period for simulations and (2) number of statistical iterations? For example, is the segmentation in the water body model application (EXAMS) unacceptably large, thus causing numerical dispersion of the chemical and, thereby, an erroneous exposure concentration. This is a discussion that we still need to address in answering this question.

Another scale issue raised by a number of panel members as well as in public comments represented by the AMEC report deals with the adequacy of the specification of the maximum radial distance and the affected populations. Our concern is that there may be facilities for which a sensitive type of receptor was not within 2 km of the source, but should not be ignored due to its importance to the human food chain pathway. In other words, the 3MRA system is not computing the total risk posed by a given site/chemical situation. The sensitivity testing necessary to address this site geometry-related issue would seem to be a high priority to maintain credibility for model results.

3. Model Comparison with Field Data

Mercury at a chlor-alkali facility

The model to data comparison at the chlor-alkali site is not expected to be very meaningful given the long (and unknowable) history of releases from the site and from other sites in the area and the fact that mercury can be transported long distances in the environment making it impossible to determine the original source of pollutant in the area. Unfortunately, the chlor-alkali facility seems to be one of the few cases where a site exists with some measurement data for both comparing the models to each other and to the natural environment. However, measurements collected at a few points in space and time give little indication of whether mercury in those samples came from the chlor-alkali site or even if those measurements are representative of what is in the environment. These data

matching exercises are questionable even when significant amounts of data are available and the pitfalls have been well covered by Oreskes et al. (1994) [Naomi Oreskes, K. Shrader-Frechette and K. Belitz (1994) "Verification, validation, and confirmation of numerical models in the earth sciences." Science. 263:641-646] (and many others over the last decade).

Validation of land-based source and watershed modules

The following comments (Thibideau) are based on the two plus pages in the section entitled "Summary of Validation" (4.2.4) in Volume 3. It is not clear what criteria are used to accept or reject the "data" vs model predictions of a particular validation test. Examples of the "moving target" criteria follow as each of the four validation activities are commented upon.

Validation by "definition" is defined on p. 4-14. I agree that using verified software components based on empirical data is an excellent approach. However, the Land-based Source Modules and the Watershed Modules each contain several of these empirical software components; but they are connected by mass balances in the hydrology model, the soil erosion model and in the constituent fate and transport model to produce the Local Watershed Model algorithm, for example. In addition, performing the mass balances requires some assumptions to be made about process structure, etc. The final result of this procedure of algorithm development includes the empirical data as imbedded elements. To claim that the final Modules are implicitly validated because they contain the imbedded empirical data is not factual. A more rigorous validation of the final module in needed in the opinion of this reviewer.

In another validation exercise for the LAU module, measured half-lives of dioxin in sewage sludge were compared. Remaining concentrations at equivalent human health risks were calculated for the LAU in order to estimate the half-lives. "The range of half-lives over the selected percentiles was 20 to 48 years, which is in reasonable agreement with the observed half-lives at several monitored sites." The numerical range is not reported; the number of monitoring sites not agreeing was not reported. This is another example of reporting the results of a validation exercise that is not fully documented and is too descriptive and subjective.

4. Model Component Comparison with Process Experimentation

Soil-column study data

The LAU Module has been compared to experimental data obtained on organic chemicals during application of municipal wastewater onto soil. Four elements of evaluation were tested: volatilization, first order chemical decay, appropriateness of the quasi-analytical solution an LAU thickness and temperature play significant role in volatilization. The volatilization rate was reported to be in the "right order of magnitude" for all categories of compounds. However, for the highly volatiles the model was consistently lower than observed. The SA with thickness showed none and the SA with temperature "showed certain sensitivity" on volatilization. The text abruptly ends on page 4-16 without comment on the other two of four elements of evaluation. This soil-column study is the nearest thing to a validation exercise for the GSCM but it appears incomplete and not well documented.

5. Model Comparison with Theoretical Expectations

The mass balance testing of the full 3MRA model framework is an excellent example of validation by comparison with theoretical expectations. The SAB panel needs a clear, unambiguous response from EPA on this task. Completing the three exercises that resulted from the fact-finding call of 18 Sept will help a lot. But even with these results in hand, we need a comprehensive response that identifies (1) where mass balance is and is not achieved by 3MRA, as well as (2) why it is or is not in each instance. The latter has to do with the limits to model system validation, and is important to users and the public. Can we frame a structured response by answering following questions?

a. Where in the model is mass balance maintained? Provide evidence that it is.

b. Where in the model is it believed to be maintained, but cannot be readily demonstrated?

c. Why is it not able to be demonstrated in these parts of the model? A statement as to the limitations

of demonstrating the mass balance, i.e. the inability to account for mass because of the types of process formulations used in the modules, the prohibitive computational burden, inter media partitioning, media-receptor partitioning, etc., is necessary and important to model validation.

d. Can the magnitude of the departure from mass balance be estimated? If so, how large is it likely to be? If not, why?

Are the I/O "wrappers" around the modules in 3MRA at all useful for computing and thus demonstrating mass balance? As I understand it, the I/O "wrappers" were designed primarily for standardization of consistent communication of inputs and outputs between the various modules in 3MRA. Is the mass being passed between modules and through the I/O "wrappers" computable? If so, can't one do a check on material balances? If not, can they be adapted to do a mass balance check? Explain.

6. 3MRA Model Comparison with Other Models

Comparison of HELP model with LAU module

A model versus model comparison was made for the LAU module against the results of the HELP model. This was a comparison of run-off and infiltration at six sites. Under the circumstances such model-to-model "benchmarking" is an appropriate validation activity. The following end-point comparisons were listed: "...on EPA expected long-term averages to be in reasonable agreement. The comparative results were mixed." "...predictions were quite similar...showed relative large differences". 'With regard to differences in infiltration...there was no bias in the 3MRA.." However for runoff the 3MRA predicted more at all sites. No numerical values were given to quantify differences. In summary the benchmarking results were ruled adequate for the 3MRA national screening-level purposes. This is another example of a lack of clear criteria for accepting or rejecting a model validation result.

Model comparison with TRIM.FaTE

The TRIM.FaTE comparison will certainly be helpful for evaluating the ability of the linked single media models to account for and balance mass transfers across media. I am very concerned about the intractability of the receptor approach that is currently used in 3MRA and depending on the outcome of the comparison with TRIM.FaTE. I will certainly have a better understanding of the potential problems or an improved comfort level with the model when this comparison is complete.

The model to model part of the comparison can potentially be very informative. To make it easier to interpret the results, the comparison should be developed on a simpler site layout using a representative range of chemicals (see figure below) and environmental conditions so that specific hypotheses can be tested. For example, the question of mass balance continues to be critical. The TRIM.FaTE model is a fully coupled model where mass is completely conserved and tracked. It can be set up using only air parcels so that the spatially averaged receptor approach using the ISCST3 can be tested independent of the other compartments. Then a simple surface layout can be added to evaluate deposition and test the assumption that "secondary transfers" are in fact insignificant. As indicated above, I also think the full model comparison as planned will be informative but I would add to that a comparison of model sensitivities for each of the estimation endpoints. This will provide a since of why the models are similar or different and point out where the models might give similar results for different reasons.

I would certainly recommend keeping metals and some pH dependent chemicals on the list as well but for the organics, it is important to have chemicals representing the four general areas of solubility parameter space. One approach that might be useful for selecting a representative set of organic chemicals is to use a figure similar to the one used by Wania (2003) [Wania, Frank. 2003. Assessing the Potential of Persistent Organic Chemicals for Long-Range Transport and Accumulation in Polar Regions. Env. Sci. Technol. 37(7): 1344 – 1351] and select chemicals that represent each of the four main regions in the solubility parameter space (i.e., air, water, solid and multimedia). I had a similar figure from some earlier work with the CalTOX model and plotted the current set of organic chemicals that are in 3MRA. The current seems to be biased towards chemicals that partition into the air. Augmenting this test set to more fully represent the parameter

space for the TRIM.FaTE comparison (and any other evaluations) is recommended.

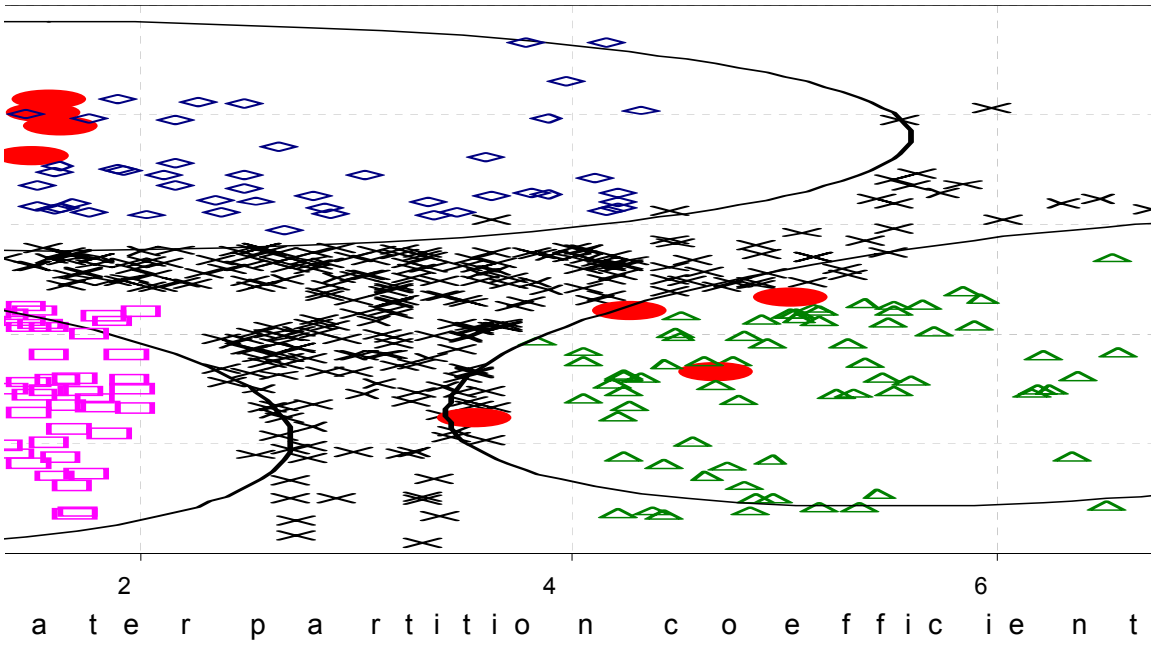

Figure 1. Example of solubility parameter space and how it influence partitioning in the environment (generated with the CalTOX model). The plot also shows the current list of 3MRA chemicals (red dots).

Additional model-model comparisons

It would also seem reasonable to request a commitment from EPA to add the results of any ongoing studies, such as the TRIM comparison with the HoltraChem studies, or MEPAS/RESRAD comparisons as Addendum Documents. Similarly, this author is aware of recent intercomparison studies of the AERMOD, AMD and ISCST3 air dispersion models. The results are providing support for a new preference of AERMOD as the "model of choice" for regulatory permitting applications for elevated sources in the near future, but AERMOD also has more complicated meteorological data requirements than either AMD or ISCST3. This may make AERMOD less appropriate for use in a screening model for ground-level sources of current interest in 3MRA.

Additionally, the British-developed AMD model is reported to have performed more precisely in a case with ground-level sources. This would suggest AMD's potential as a future "improved" model for 3MRA, unless the charter for 3MRA expands to include elevated sources. This air model situation gives but one example of how the protocol for using 3MRA will need to remain somewhat flexible to accommodate new module developments; but the EPA has stressed that has already been considered in their development of a standardized set of interfaces for each of the current modules. This is an operational plus, but the substitution of any newer model will remain a 'model validation' challenge.

These are but two examples of the projects supporting regular upgrading of constituent sub-models ("modules") used in regulatory applications. It would also seem reasonable to discuss them under the subject of "future upgrades" to 3MRA, even though the current sensitivity testing is still being completed. Similar discussions might also include the related new model developments concerning (1) uptake factors in food-chain models, (2) use of ecotoxicity benchmarks, and (3) special models needed for simulating mercury behavior in the environment. Improvements like these are continually arriving from other model development and testing projects within ORD's realm. It may increase the confidence of the public that the 3MRA system is "scientifically sound" if this principal of continuing improvement is given more emphasis in each volume of the 3MRA documentation.

7. Peer Review and Resulting Revisions

Peer review is one of the key activities recommended by Beck et al. (1997) that EPA has included in their evaluation plan to insure that the 3MRA is scientifically sound. The level of peer review that the individual model components received is commendable and the individual comments (in the few reports that I looked through) seem very useful. However, model changes or improvements that resulted from the peer review are difficult to track. In a number of cases it appears that the comments were dismissed without a stated reason.

For example: in the peer review of the ISCST3 module, concern was raised about the use of a single scavenging coefficient for all gaseous chemicals as indicated in the following excerpt from page 11 of the air_sum.pdf file posted on the 3MRA website -

> "2. Is the use of the small particle assumption in place of chemical specific scavenging coefficients appropriate for this assessment?
> Dr. Hanna: This seems like a poorly-thought-out assumption. Has anyone checked this out with some calculations using these alternate ways of treating the topic of scavenging? I can think of some obvious problems, since there are large differences in the scavenging coefficients for gases, depending on the chemical properties of the gases (solubility in water, Henry's constant, presence of other chemicals, etc.) and the characteristics of the atmosphere and the surface. This assumption should be reviewed by scavenging experts such as George Slinn."

Yet, in the discussion of how wet deposition was incorporated into the ISCST3 module (section 4.3.4.3), the Agency makes the following statement –

> "To reduce the number of model runs required for a 3MRA application, EPA configured the Air Module to use a single vapor-phase scavenging coefficient value for all contaminants that causes them to be scavenged as if they were small particles."

Indicating that no change was made to address the reviewer's comment. I actually agree that assuming that gaseous chemicals with a wide range of air/water partition coefficients will all behave the same during a rain event lacks scientific credibility. However, there may be a good reason for this modeling simplification (run time?) but I would like to be able to go to a summary table that explains the reason so that I, as a reviewer or as model user, can know that it was considered and make my own judgment about the merits of the decision.

In the agricultural food-chain module, one reviewer expressed concern that only dissolved phase contaminant in water was considered in the exposure calculation for farm animals while "cows consume dissolved and particle bound contaminants" (bottom of page 12 in the peer review document on the 3MRA web site) – yet the module still excludes suspended particles from the ingestion pathway and no explanation is provided as to why the review comment was dismissed. The same reviewer went on to say that:

> "In my view biotransfer factors are a very crude way of estimating uptake. There has been a lot of work done on uptake of chemicals by farm animals and models have been developed by workers such as McLachlan, Sweetman and others, mostly in Europe.

> This work has been totally ignored in favor of flawed correlations. Serious consideration should be given to scrapping the entire biotransfer factor approach. A more honest approach for this program would be to state that given the present state of the art, these general calculations can not be done with sufficient accuracy to justify their present inclusion. …"
>
> Did this comment have any effect on the final product?

Granted, many of the comments that come out of a peer review are difficult to address (and often contradict comments on the same question by other reviewers) but if we are to accept that the peer review process succeeded in its stated goal (i.e., "to ensure that the theoretical concepts describing the processes within the release, fate, transport, uptake, exposure, and risk components were adequate representations of the processes to be evaluated." (quoted from 3MRAVol3_02 page 1)), then the link between the peer review and the final product needs to be made.

The panel recommends that the peer review should be completed. To accomplish this part of the model validation the Agency needs to provide a detailed response (or at least a list of summary responses) to the peer reviews performed on the 3MRA modules, including a description of changes that were made to the model and/or the rational for dismissing the reviewer's comment.

8. Evaluation of Model Forecasting Simulations

Selection of representative sites

The selection process for identifying the 201 representative sites from the national survey is key to the question of whether the 3MRA will produce results that are useful for the national application. The idea of using a subset of sites as surrogates for all possible sites is well accepted and scientifically defensible but I question whether a simple random sample (~ 7% of the total number of sites (i.e., 201 sites selected from >2800)) is really representative of the national population of sites. What seems to be missing is an explanation of where the sample size (i.e., n = 201) came from and/or whether the resulting sample was stratified appropriately on the variables of interest (WMU type on site, eco-region, meteorology, population density, number of farms, hydrogeology, topography …). Stratified random sampling is certainly a well-established technique, and the use of stratified sampling guarantees that the resulting sample has the correct population proportion of each variable of interest. The 201 sites may in fact be

representative of all industrial non-hazardous waste management across the U.S.; but this assertion needs to be demonstrated at some point in the documentation.

The panel therefore makes the following recommendation relative to national representative site selection: Identify the important characteristics that the Agency is trying to capture in their data set and plot the proportion of the 201 sites having each of these characteristic along with the proportion of the full sample (>2080 sites) having the same characteristic and, if there is a difference, explain why that difference was deemed acceptable. For example, if a certain percentage of sites in the national survey were near urban areas then a similar proportion of sites near urban areas should emerge in the sampling of 201 sites. The list of site-based data might be a good starting place for important characteristics of the sites.

Sensitivity analysis

For a model of this complexity, I (Maddelena) would put as much importance on the ability to perform an informative sensitivity analysis for each application as I would place on separating uncertainty and variability or even on the uncertainty analysis itself. The 3MRA system does not seem to be capable of executing a sensitivity analysis at this time. Much discussion is provided on the different options and methods that are available for the application but none has been implemented to date. The question that this raises is whether the model is even amenable to SA on the level that is needed to understand the results.

Hodges and Dewar (1992) [James S. Hodges and James A. Dewar, Is it you or your model talking? A framework for model validation. Prepared for the United States Air Force, United States Army and Office of the Secretary of Defense by RAND. Report # R-4114-AF/A/OSD. 1992] suggest that when a model is used to make a prediction then that prediction must be accompanied by a statement about its accuracy (i.e., the model says X „ Y) and an argument for why someone else should believe that statement. Uncertainty analysis and validation lead to the statement of accuracy („ Y) and validation alone supports the statement of believability. It is clear that the Agency has accepted the fact that 3MRA cannot be validated so the statement of believability will need to be based on something else. In this case, that "something else" includes the quality of the data and modules (internal constituents of the model according to Beck et al (1997)).

The EPA has put a lot of effort into evaluating the quality of the input data and the individual modules in the 3MRA, with varying degrees of success. It is clear that not all of the modeling components are associated with the same level of scientific credibility. In other words, after considering all the effort that went into Volume 3, I have a lot more confidence in the data than I do in the modules and I have a lot more confidence in the EPIC module than I do in the farm food chain module... Given that all the components of the modeling suite are not associated with the same degree of credibility, it is critical that the modeler identify what components (i.e., data, modules, and imbedded assumptions) are most important for any given application. This is done using sensitivity analysis methods.

Because we are now trying to use a single multi-media modeling framework for all chemicals, even those that are predominantly single-media pollutants, understanding the sensitivity of 3MRA to its inputs might be one of the biggest challenges that the model developers face. The list of important modeling components will differ for each modeling scenario (chemical, site, population and possibly time). (I can provide reference for this statement if needed). Therefore, we cannot expect to perform one mega sensitivity analysis using the SuperMUSE and make a conclusive statement of why the model behaves in a certain way. Rather, each model outcome or prediction must be associated with a scenario specific sensitivity analysis and an interpretation of the results of that sensitivity analysis.

The idea of relating the quality of the model to the relative numbers of "key" and "redundant" model parameters (i.e., the Sensitivity-Based Performance Validation) is not helpful in the context of a model that is designed to work with a wide range of different pollutants. The more "multimedia" a chemical behaves the less redundancy one would expect to see in the inputs to a multimedia model. I would expect to see a lot more redundant parameters when a multimedia model is run for vinyl chloride (VC) than for hexachlorobenzene (HCB) but that doesn't mean the model will work better for HCB. It simply means that the fate of VC depends on a much smaller number of fate processes and interacts with a smaller number of environmental media than HCB. In fact, for this very reason, I would expect the model to

provide a more believable prediction for VC than HCB even though the degree of redundancy is likely to be much higher for VC.

The panel suggests the following: Finish developing the sensitivity analysis application in the 3MRA so that each outcome from a particular "National Assessment" can be presented along with a list of important model components for that particular outcome of that particular application. The presentation of sensitivity should enable the user to clearly identify (1) the important exposure pathways, (2) the important modules and (3) the important inputs. Given the complexity of the model, it might be best for the Agency focus resources on getting a single SA method working and then expand to some of the other methods described in the report as resources permit. The important thing is to get the full sensitivity analysis working. Without the sensitivity analysis the panel maintains that the results are not as useful or acceptable as they could be. Given that Monte Carlo analyses are planned already, the simple rank correlation approach to estimating sensitivity might be the best place to start.

Challenges that will need to be addressed during this analysis include:

1. The importance of a specific input will likely change from site to site depending on environmental conditions and from location to location at a given site depending on where the chemical accumulates and this will need to be captured and interpreted in the sensitivity analysis results.

2. Importance of a specific input is likely to change over time (from year to year) as the chemical accumulates in different media and migrates towards the target and this change will need to be captured and interpreted in the sensitivity analysis results.

3. Any informative SA approach is likely to be extremely data intensive for this type of model. It might be necessary to make some choices about outcome of interest up front so that data storage can be managed.

4. All of the sensitivity analysis methods described in Volume 4 require that the model component that is being evaluated be stochastic, or at least be changed from run to run. Of the 900+ inputs currently in the model, all of those that are treated as "constant" will be excluded from the SA. In addition, the importance of imbedded assumptions will also be missed. For example: The assumption that BCFmilk and BCFbeef are 1 if the physchem properties

fall outside the range used to develop an empirical relationship will not be tested. I think there are a number of these constant and or imbedded assumptions in the various modules and these will need to be included in the SA.

Validation at the regulatory decision level

Given the complexity and very broad scope of the 3MRA framework and the resulting great difficulty in fully validating the model prior to using it as a management tool, the panel is wondering whether a better question to ask for this type of model might be whether the use of the model leads to "correct" decisions. Given the long history of RCRA and Subtitle C (> 25 years), it may be possible to pose questions where the answers are actually knowable. For example, are there chemicals that under no circumstances should they be allowed to exit Subtitle C? Are there chemicals where a consensus on exit levels has been reached through some other process? Are there chemicals that clearly should be allowed to exit? If so, then it would be relatively straight forward to model decision level performance by setting up a semi-blind study where only the physical chemical properties of the constituent in the waste stream is known and the 3MRA is used to come up with recommended exit levels.

This type of comparison can determine if the model outcome is biased high, low or random. If the outcome is random then we can start to evaluate the actual range of this random error. Depending on how many "knowable" outcomes exist, one could begin to construct a statement about the likelihood that the model will correctly (or incorrectly) allow a contaminant to exit Subtitle C. Although it has not reached the archive literature yet, there is actually some precedent for this approach now in the persistence and long-range transport (P&LRT) modeling community. A dataset has been developed with several hundred chemicals where a number of them are identified as definitely subject to LRT while others are definitely not. This work is going on right now but if there is interest I (Maddelena) can try and find out more details on the approach.

Such an exercise would certainly extend the method proposed by Beck et al (1997) without requiring additional environmental monitoring data or even worrying about the performance of the model at individual sites. It may be beyond the scope of this project, but the Agency should explore the options for compiling a list of chemicals for which consensus has already been reached

as to whether the constituent should be allowed to exit or not and if allowed, at what level and then compare these decisions with the outcome of the 3MRA national simulation for those chemicals.

Summary

A large number of different 3MRA model validation exercises were reported in the documentation. These included all of the above performance indicator categories. In general, many of the reported validation outcomes were reported using qualitative descriptors that made it very difficult for reviewers to assess the true success of a given test. This general failing must be dealt with before a more informed assessment of the 3MRA validation can be made.

Some validation exercises were applied to both individual components of the overall system and to the overall system itself, while others seemed to focus on modules more than the whole system. The panel believes that: the science and resulting process formulations in individual models should receive more attention; and the performance of the whole integrated system relative to its stated purpose should receive more attention.

It is clear from what has been reported thus far that the 3MRA modeling system has not been validated in the conventional sense of process-oriented environmental fate, transport and risk models. It is very important for EPA to continue to pursue site-specific model comparison with measured data, site-specific comparison with other multi-media models, and continued development of the SA/UA as part of the model confirmation approach. Unless and until these ongoing activities can be integrated into the confirmation process, the utility of the 3MRA system is speculative and it may be "premature" for the panel to definitively answer this question. Indeed, models and sub-models are continually evolving and improving as we expand our knowledge base and acquire new data. It is admirable that EPA has designed this modeling system to be easily revised and upgraded. In order to move forward in the absence of a complete and formal model validation, it might be worth considering an adaptive management approach whereby continued follow-up observation of actual sites can be used to reduce model uncertainty and thereby improve the decision process.

This October 7 draft report has been prepared for discussion at the 3MRA Panel's October 9, 2003 Conference Call Meeting. The Panel has not reviewed this draft.. Therefore, this draft cannot be said to represent the consensus of the Panel.

Charge Question 4: Is the documentation for the 3MRA Modeling System adequately designed and prepared? Does the SAB have additional suggestions for improving the presentation of the comprehensive set of materials related to this modeling system?

General Comments

Based on its review to date of the 3MRA documentation (five volumes), the SAB Panel finds the documentation to be reasonably well presented in a reasonably well organized fashion, however the documentation requires important revisions before it can be considered adequate. Several panelists with familiarity with the earlier 1995 HWIR documentation that preceded 3MRA indicate that the 3MRA documentation is a significant improvement over the HWIR materials prepared in 1995. It seems clear that many earlier criticisms about the clarity and completeness of the deficiencies in the HWIR documentation have been taken as constructive criticism by the EPA authors.

Given the challenging volume of material included in the 3MRA modeling system, it is generally readable if taken in modest doses. The level of detail provided helps the reader to understand both the strategic thinking that went into its planning, development, verification and (limited) validation testing. The organization of the five volumes, with detailed tables of contents make it relatively easy to limit reading the to subjects of greatest concern. For the reader who is deeply interested in the model framework, the development and verification history, or the specific modeling algorithms used in the 17 simulation models, the documentation is reasonably well designed.

The Panel comments to date point to several areas that require clarification, and possibly significant revision, in order to make the 3MRA documentation clear, transparent, and more understandable:

1. There is a need to develop a more "digestible" summary that describes the 3MRA in layman's terms. The sheer volume of material, combined with technical jargon covering many disciplines, makes for a "dense" read for a non-technical audience.

2. The discussion of uncertainty, variability, and sensitivity concepts relating to the Monte Carlo analysis (Volume IV) reads too much like an academic treatise, and fails to explain what was done to address variability and uncertainty in the 3MRA. The explanations in this volume require significant revision in order to make the actual Monte Carlo implementation of the 3MRA understandable and transparent.

3. The number of operational and input parameters that go into the 3MRA is so large, that it is at present almost impossible for someone not fully versed in the model details to grasp which are based on data, and which are operational assumptions. It appears there are key variables that are based on operational assumptions (an example is the "fraction hazardous waste" term), and it is crucial that EPA clearly summarize these more concisely, and provide a more intuitive understanding of how they influence the model formulation.

The remainder of the Panel comments on the 3MRA documentation are provided as "specific comments" below. Those that are not specific to a particular volume are presented first, followed by comments that are directed toward specific volumes. Finally, while it is not the Panel's intent to review the document in terms of style, grammar, or typographical issues, to the extent we have input, these comments are noted as "nits" at the conclusion of these specific comments.

Specific Comments (Not Specific to Individual Volumes)

- The documentation indicates in numerous places that the earlier problems of mass balance violations in the 1995 predecessor models have been corrected. It will be important to provide the user with access to sufficient model outputs (source terms, mass fluxes, mass in environmental compartments) to allow the user to confirm the preservation of mass balance between modules in a transparent manner.

- The 3MRA modeling system should have the option to print sufficient model details (to an output file) regarding the combination of major variables that yield specific outcomes for the Monte Carlo simulations in order to examine the combination of input variables that lead to various exposure and risk outcomes.

- A glossary should be included. Many words in the documentation are not in common use or are defined in this context differently from their everyday use. Perhaps EPA could consider creating a searchable electronic index of

the entire documentation where a modeling practitioner could query a topic and the "help" module would identify a list of choices within the documentation from which an individual could find details on that topic.

- A 70-year human exposure window that takes population weighted averages of the various cohorts should be scanned over the 10,000-year time horizon in order to capture the lifetime average daily dose estimate needed to calculate risk. The available documentation appears to indicate that point values of risk are selected by the present version of 3MRA.

Volume I

Volume I provides a useful overview of the purpose and structure of 3MRA, and prepares the reader for the information provided in the subsequent volumes. The repetition of fundamental information is generally summarized adequately in each subsequent volume so that generally it could be read independently of the other volumes. There remains a need to develop a more "digestible" summary document, aimed at a non-technical audience.

Although Volume I goes part way toward fulfilling this need, it needs an expanded executive summary written in layman's terms. EPA could consider developing a more graphical summary (possibly using animation) as one means of more effectively conveying the complex topics addressed in 3MRA. Lacking a more understandable expanded summary, the only people who are likely to understand the system truly will be the developers. If that is the outcome, the decision-makers will not relegate their power to the creators of the model, and the technical tools will sit on the shelf and collect dust.

Competing with the need for a more widely understandable summary, is the need to add to the summary in Volume I to provide sufficient detail for the more technically oriented reader. Thus, Volume I would benefit from the addition of some the additional technical information in Volume II and Volume IV in order to provide sufficient information on the intended strategy of application and the interpretaion of modeling results, especially the uncertainties.

- A clearer description of what the exit concentrations refer to is required. The Panel's understanding is that the exit level represents the concentration in the waste stream as it would enter a WMU, and not the concentration of the contaminant within the WMU (although there remains ambiguity on the source term as noted by other Panel comments elsewhere). It seems fundamental, but because of the terminology used (exit levels), it warrants supplemental description. Another area that where additional explanation would help the reader, is reiterating the fact that that the exit concentrations are chemical-specific and are for single, or limited WMU combinations.

- The document should be clear as to whether exit levels are being calculated for a "site" as a whole, versus WMU-specific exit levels. For example, on page 3_4 in Volume IV it is suggested that EPA is considering developing WMU-specific exit levels (which the Panel believes has merit). Yet in Volume I and elsewhere the exit levels appear to be defined only as the result of the sum total of all combined WMUs at a particular site. The document should be clear on this point one way or another.

- The concept of risk bins (intervals) is a new one and one that needs further explanation. In addition, the notion of "percent of population protected" requires further clarification. As the phrase is stated, it suggests that the 3MRA risk results are multiplied by the size of a specific population at a site to determine what fraction/percent of that particular population would exceed a specified risk level. Yet, it is the Panel's understanding that the 3MRA risk outcomes are not multiplied by the size of the (known) population within 2km of the modeled sites. More clarity is needed to explain how the risk calculations are translated into a "percent population protected."

- Some panel members comment that Volume I uses graphics especially effectively to orient the reader as to model

structure and function, although others felt improvements are needed.

- The equation notation used in Volume I makes it easier for a modeler who specializes in one topic area to see the clear linkage to related topic areas and algorithms, however there are instances where different notation is used for the same equations depending on specific sub-model considerations which does create the possibility of confusion and ambiguity.

- Volume 1 (and elsewhere) could provide more context to the reader on the nature of the 201 sites in the database (by region, size, industry, WMU and waste types, etc.). Although there are indeed 201 sites in the database, exit levels for many solid wastes will be set based on specific land application units (LAU), of which there are only 28, landfills (only 56 in the database), etc. Likewise, only 137 sites managed liquid wastes, so exit levels for liquids will be based on only 137 sites. This should be clarified in the documentation. As it stands, the indication that the database included data for 201 Subtitle D sites can be erroneously interpreted to imply that the site database is more robust than is in fact the case for specific WMU types.

- Vol. 1, Fig. 1-2. This figure (which also appears elsewhere) is very busy with a multitude of interconnected compartments such that its value to the reader becomes lost. Figure 2-3, which has similar elements, is much more intuitive. In addition, Figure 2_3 could possibly be enhanced with the addition of the model(s) that are associated with each module (where appropriate and without adding undue clutter to the figure).

- Vol. 1, Fig. 1-4 would be more helpful if the Y_axis were changed to depict the probability of protection rather than the probability of non-protection. Both the Y-axis, and X-axis of this plot require better labeling to clarify them (this applies to many of the plots depicting the risk outcomes in the form of probability curves).

- Vol. 1, Fig. 1-5 isn't clearly labeled and seems counterintuitive. Is the Y axis % sites and the various curves % receptors at the sites or vice versa? Either way, for a given waste concentration these percentages should move inversely, i.e. a high percentage of sites might achieve 50% protection but only a few would achieve 95% protection (this is shown in Fig 1-6 and 1-7). More importantly, the Panel is not convinced that the 3MRA uncertainty analysis actually accomplishes a "pseudo-2D" analysis, and unless the pseudo-2D analysis can be shown to provide a meaningful analysis of uncertainty (separate from variability), these figures will require revision to remove the notion of "2D" analysis.

- According to Vol. 1, page 5-6 bullet 2: "concentration can be adjusted for other wastes which do not contain the constituent." Again on page 9_7 of Volume 1 there is a reference to incorporating a "fracture multiplier" for the aquifer module. It would be helpful to add a table to the documentation showing all the options and ad hoc adjustments such as these that are contained within the 3MRA model and which option(s) are selected for the purpose of setting national exit criteria.

- Inclusion of a bio-uptake factor in the human and ecological exposure modules would enhance the versatility of the model. Even if the default value were one, the inclusion of such a parameter would provide a concrete basis for a future site-specific analysis employing waste-specific bio-uptake data.

Volume II

- Although the data sets used are generally identified, it would be helpful to provide a concise summary (perhaps by module) of the date, size and scope of the data set, and other important contextual information that identify the major data sets used to support the models. While this information may exist in the voluminous documentation, a concise summary in one location would be helpful if at all possible.

- There are many data distributions that are indicated as being selected using "best professional judgment." Again, it would be useful to summarize in a more "global" manner, the types of important model parameters that are based on empirical data, and those that are based on professional judgment.

Volume III

- Volume 3: From a readability standpoint, Chapters 4 and 5 should be switched, i.e. "Verification and Validation of 3MRA Site-Based Data Collection and Processing" (Chapter 5 Volume 3) should be presented before "Evaluating Quality of the 3MRA Modeling System Modules" (Chapter 4 Volume 3).

Volume IV

To characterize and bound the uncertainties for the policy marker is essential for them to understand the potential impact that their decision will have. It is equally important for them to understand how to delineate that uncertainty and comprehend how sensitive the 3MRA system is in its yielding exit concentrations. Therefore, it is important that the material in Volume 4, Uncertainty and Sensitivity Analysis, be either re-written at a more understandable level or take the majority of the material and place it in an appendix for the reader/user to pursue at a his or her leisure. The chapter is dense, even for one whose vocation is risk assessment.

Section 2 in particular reads like a textbook in some places. The panel suggests that the discussion be more focused on the actual methods used in the 3MRA and how the results thereof should be interpreted for decision-makers and stakeholders. There simply is too much tutorial information that gets in the way of learning what uncertainty, variability and sensitivity analysis is all about in 3MRA. The need for clarity and simplicity of explaining how 3MRA addresses uncertainty versus variability (if indeed it does this explicitly) takes precedence over completeness in describing the "taxonomy" of sensitivity, uncertainty, and the like. The document should target the model user as the principal reader, not the academic scholar. As it stands, the document is guilty of swamping the reader in a "sea of linguistic ambiguity," (e.g., p. 2_13) rather than providing clarity.

There is a rich and often confusing lexicon of terms describing uncertainty, variability, and sensitivity in the literature, and the report devotes a lot of ink to reviewing this literature. However, it is not the authors' primary job to sort all that out for the benefit of the 3MRA user. Rather their responsibility is to clearly and unambiguously define how they are using the terms uncertainty, variability, and sensitivity (and their derivatives) in the 3MRA context. What do these terms mean, how are they defined for 3MRA? Specific examples should be provided, rather than speaking in vague generalities. How are they estimated, examined, analyzed, and interpreted in 3MRA? Only then should the authors elaborate on how their use of these terms/concepts/analyses, etc. relate to others in the literature, and only as is necessary to clarify for the reader/user what 3MRA is doing. Furthermore these elaborations can be relegated to an appendix .

The documentation must be consistent in its treatment of variability and uncertainty. Although the documentation (e.g. Section 2.6) spells out the various kinds of uncertainty and identifies those that the model addresses and those that it does not address, other places might give the impression that variability and uncertainty are separately quantified. Volume 1, Section 1.2.1 states, "Quantifying variability and uncertainty in exposure and risk estimates is an important capability of any modeling system. The 3MRA modeling system was designed with a two-stage Monte Carlo analysis capability, which enables users to distinguish between variability and uncertainty in input variables". Section 2.1.1 (page 2-4, paragraph 1,) states "the distilled output prediction can, for example, be represented as predicting 90% receptor population protection at 95% of sites with a 98% probability (or confidence or belief) of meeting this 'dual criteria' population protection level." The Panel does not believe that this quantitative separation of variability and uncertainty has been achieved in the current modeling system, nor does it believe that a more rigorous 2-Stage Monte Carlo analysis that would separate and quantify uncertainty is realistic in the near future. The documentation should be clear and consistent on this point. Although the panel believes that uncertainty has not been and cannot be quantitatively characterized at present, we recommend that it be addressed qualitatively, noting that wherever variability is quantified, some unknown fraction of what is being called variability may actually be uncertainty.

¶ Statements in Volume 4 mention that the reader might get conflicting impressions as to whether 3MRA ver 1.0 actually distinguishes between uncertainty and variability. The Panel agrees that discrepancies between statements in Volume IV and Volume I do indeed confuse this issue, and these discrepancies should be resolved.

¶ In addition, the document creates confusion in the reader regarding the various versions of 3MRA (e.g., ver 1.0, and ver 1.X, ver 2.0). The additional functional capabilities of 1.x and 2.0 over 1.0 are outlined in the report, but what about problem solving? What kinds of problems can the panel investigate with the PC version distributed to us? Please provide some scenarios. On the one hand it appears that the pseudo_2D analysis requires the SUPERMUSE, yet this is not completely clear from the documentation. The document should be very clear on what the distinctions are between the versions, and which version(s) is/are being used to develop exit levels.

¶ In the Monte Carlo analysis, toxicity parameters are treated as fixed when, in fact, they are both variable (not everyone's threshold is the same) and uncertain (most criteria are based on laboratory animal data). This has the effect of artificially narrowing the distribution of risk. In addition, because the fixed values are upper-end estimates, the distribution of risk versus probability is artificially shifted to the right (meaning that a given scenario appears more risky than it really is). Ideally, toxicity parameters should be entered as distributions, like other variable and/or uncertain parameters. It should be a long-term goal of EPA to develop distributions for toxicity parameters. However this is clearly not going to happen in the time frame needed for the current version of 3MRA. So at the least, the documentation should make it clear that the risk and hazard estimates corresponding to the exit levels are exaggerated. [Note: this issue is not so much a "documentation" issue as a fundamental construct of the 3MRA. This issue should probably be raised elsewhere.]

¶ Software issues/Initial Conditions. Some concentration ranges need to be expanded to spread out the

probabilities. It does little good to have 100% protection at all concentrations.

- The ongoing plans for further Monte Carlo model sensitivity and verification testing could be improved/clarified. It included many development program details, such as budget estimates and schedule timing that, while they may be of interest to some readers, seemed peripheral to the mission of the main document. They may just be an indication of a "work in progress", but those facts relevant to the more permanent readership could be included as an Appendix or Addendum.

- The summary of model parameters in tables in Section 8 (e.g., Tables 8-9a, b, …) should include the 2nd moment (e.g., variance or standard deviation) where appropriate when describing probably distributions. Currently, only the first moment is provided, with a range.

Volume V

- In attempting to run the model and its initial example cases, some Panel members found that information from both Volumes IV and V contained needed model summary material and descriptions of application methods before the model could be run, but the information could be improved by including it in a single volume. A set of several sections seemed to contain sufficient information for someone with a general knowledge of the purpose of the model and its constituent elements, but who wanted to run the model with minimum time devoted to "refresher" reading. A candidate outline of the material that would go into such a 3MRA User's Manual is attached. The outline draws information from Volumes IV and V, and leaves Volumes I, II and III for a separate reading exercise.

- It is not till Chapter 4 that the reader finds the section on "Installing the 3MRA Modeling System", i.e., "getting started." This section should be up front, with the current preliminaries relegated to inferior positions in the document. After all, much of the current preliminary sections are recapitulations of material in the

other volumes. They are not primary to the user manual and example simulations.

- Because one of the major requirements of the system is to implement on IBM-compatible personal computers (thereby making it an accessible PC-based system), it would be useful to present the minimum requirements upfront. In fact, it would be especially useful for hose who are not technically adept, nor have up to date systems not only in the User's Manual but also in the very beginning (maybe as a separate stand alone box). Furthermore, the minimum requirements as stated (64 megabytes of RAM) appear to be incorrect. Some Panel members systems could run a portion of the program and then simply could not continue because it didn't have enough "horsepower".

- The documentation in Volume V, Section 4.3, devoted to Post Simulation Analysis is a candidate for further improvement. The authors may have assumed that the typical reader of Section 3, particularly Section 3.3.9 would have a reliable memory of how the model output was organized and how all of the postprocessors use those files. The current documentation was a bit abbreviated and could lead to new model-user frustration, but with modest user training, could not be greatly faulted.

- The sub-heading of "Consolidation of Risk Time Output Data" in Section 2.0 of Volume 5 seems out of place. Would it better identified as 2.1.3.2, or because of the importance of the content it is conveying, would it better suited to be the fourth part of the description of the conceptual modeling approach and labeled 2.1.4. This section would really be enhanced with a graphic displaying how consolidation of data occurs.

- Additional examples and model scenarios. Additional simulation exercises (some example problems) in Volume 5 might be useful. For instance:

  - What happens to the base simulation if you change X? Then explain what changes and why it does.\

How do I do 2 simulations, rather than the 1 described in the manual?

What happens if I change the random number seed? The output changes but what does this change in the output mean? What feature of 3MRA does a change in random number seed reflect?

3MRA - Q4: Candidate Outline for Improved "3MRA User's Manual"

.1 INTRODUCTION [Combination of present IV (I.0) and V (1.0)]
.2 OVERVIEW OF SCIENCE [Current IV (3)]
.3 OVERVIEW OF 3MRA VERSION 1.0 [Current IV (4)]
.4 MODEL METHODOLOGY SUMMARY [Current V (2)]
.5 INSTALLATION AND USE OF 3MRA [Current V (4)]
.6 CASE EXAMPLES
1 Single Site, Single Realization [Current IV (3.2)]
2 Example Benzene Case [Current IV (7)]
3 Example Mercury Case [New Example from model validation experience]
.7 INTERPRETATION OF RESULTS AND UNCERTAINTIES [Current IV (1.3, 7.2)]
.8 REFERENCES
.9 TECHNICAL SUPPORT FOR CURRENT AND FUTURE 3MRA APPLICATIONS

Appendix A – 3MRA Technology [Current V (3)]
Appendix B – 3MRA Inputs & Outputs [Current IV (8)]
Appendix C - Probability Models and UASA Applications for 3MRA [current IV (2)]
Appendix D – 3MRA Version 1.X Enhancements [Current IV (6)
Appendix E – The Supermuse System for Testing 3MRA [Current IV (5)]
Appendix F – UASA Plan [Current IV (9)]

Much of the inspiration for this approach came from trying to run the model the first two times. Because the initial attempt immediately followed a reading of all of Vol IV, including Section 3, as well as Volume V, the logic seemed relatively clear. However on subsequent return, it seemed difficult to remember where some of the key instruction material was located: Volume IV or Volume V?.

"3MRA Documentation Nits"

- Throughout the document, reference is made to "soil concentration," "air concentration," etc. While it may seem cumbersome, it is more appropriate and correct to refer to "chemical concentration in soil," chemical concentration in air," etc.

- The word "data" is plural. There is not a consistent treatment of the verb form that follows data.

- There were occasions when tables and figures referenced in the text were either not present, or incorrectly referenced (see Volume V, p. 2-3 and p. 2-15 as examples).

- There are occasions where the notation used in figures differs from the notation used for variables in the text (capitalization, acronyms, etc.).

- Vol. 1, p. 5-14. The boundary condition in the second bullet appears to be inconsistent with the statement in the bullet on the bottom of p. 5_24.

- Vol. 1, Section 5. There are examples (e.g., equation 5-6 and 5-16) where the governing equations appear to be presented using somewhat different notation. The notation used for a given module component should adopt the consistent notation, and avoid introducing unnecessary "variants" to the equations, unless there is a compelling reason for alternative formulations.

- Each volume is a standalone document, therefore it would be helpful for either a header or footer that contains a reference to what volume it is.

- A more judicious use of commas would enhance the overall reading, especially for those chapters whose writers prefer to use long sentences.